ISSN 1684 – 8403
Journal of Statistics
Vol: 13, No.1 (2006)

A Class of Selection Procedure for Unequal Probability Sampling Without Replacement and A Sample of Size 2

Abdul Basit[1], Muhammad Qaiser Shahbaz[1]

Abstract

A new selection procedure for unequal probability sampling without replacement and a sample of size 2 has been obtained. Some results regarding probability of inclusion and joint probability of inclusion has been verified special cases has also been obtained. Empirical study has also been carried out.

Key Words

Unequal Probability Sampling, Horvitz–Thompson estimator, Brewer procedure, Variance estimators.

1. Introduction

Sampling with unequal probabilities has its origin way back in early forties. The first theoretical framework of unequal probability sampling has been given by Hansen and Hurwitz (1943) when they proposed their estimator of population total for use with unequal probability sampling with replacement. The estimator proposed by them is

$$y'_{HH} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{p_i} \quad , \qquad (1.1)$$

where p_i is probability of selection of i-th unit.

[1] Department of Statistics, GC University, Lahore, PAKISTAN
Received Jan. 2007, Revised April 2007

The concept of unequal probability sampling without replacement was first introduced by Madow (1949) but no theoretical framework was given. Horvitz and Thompson (1952) were the first to give theoretical framework of unequal probability sampling without replacement. They also proposed their selection procedure and an estimator of population total. The estimator proposed by Horvitz and Thompson (1952) is given as:

$$y'_{HT} = \sum_{i\in s} \frac{Y_i}{\pi_i} , \tag{1.2}$$

where π_i is probability of inclusion of i-th unit in the sample.

Horvitz and Thompson gave following variance formula for estimator (1.2).

$$V(y'_{HT}) = \sum_{i=1}^{N} \frac{(1-\pi_i)}{\pi_i} Y_i^2 + \sum_{\substack{i,j=1 \\ j\neq i}}^{N}\sum \frac{(\pi_{ij} - \pi_i\pi_j)}{\pi_i\pi_j} Y_i Y_j \tag{1.3}$$

an alternative expression, for fixed n, given by Sen (1953) and independently by Yates and Grundy (1953), is:

$$V(y'_{HT}) = \sum\sum_{j>i}^{N} (\pi_i\pi_j - \pi_{ij}) \left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j} \right)^2 \tag{1.4}$$

An unbiased estimator for variance given in (1.3) was proposed by Horvitz and Thompson. The estimator is given as:

$$var_{HT}(y'_{HT}) = \sum_{i=1}^{n} \frac{1-\pi_i}{\pi_i^2} y_i^2 + \sum_{\substack{i=1 \\ j\neq i}}^{n}\sum_{j=1} \frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j\pi_{ij}} y_i y_j \tag{1.5}$$

The variance estimator given in (1.5) is unbiased for variance given in (1.3) but it may assume negative values for some of the pairs. An unbiased estimator for variance given in (1.4) was proposed by Sen (1953) and independently by Yates and Grundy (1953). The estimator is given as:

$$var_{SYG}(y'_{HT}) = \frac{1}{2} \sum_{\substack{i=1 \\ j\neq i}}^{n}\sum_{j=1} \frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 \tag{1.6}$$

Since the emergence of Horvitz–Thompson (1952) estimator, number of selection procedures have been developed that can be used with this estimator. For a comprehensive review of these selection procedures one can see Brewer and Hanif (1983) and Chaudhry and Vas (1986). In the following section we have given a new selection procedure for sample of size 2.

2. New Selection Procedure for n = 2

In this section we have given a new selection procedure for use with the Horvitz–Thompson (1952) estimator and a sample of size 2. The new selection procedure is given as:

- Select first unit with probability proportional to $p_i^{\alpha}\left(1-p_i^{\beta}\right)/\left(1-2\,p_i^{\beta}\right)$ and without replacement
- Select second unit with probability proportional to size of remaining units

The probability of inclusion for the i-th unit in the sample for this selection procedure is given as:

$$\pi_i = \frac{1}{\sum_{i=1}^{N} p_i^{\alpha}\left(1-p_i^{\beta}\right)/\left(1-2p_i^{\beta}\right)}\left[\frac{p_i^{\alpha}\left(1-p_i^{\beta}\right)\left(1-2p_i\right)}{\left(1-p_i\right)\left(1-2p_i^{\beta}\right)} + p_i\sum_{j=1}^{N}\frac{p_j^{\alpha}\left(1-p_j^{\beta}\right)}{\left(1-2p_j^{\beta}\right)\left(1-p_j\right)}\right] \quad (2.1)$$

The joint probability of inclusion for i-th and j-th units in the sample for this selection procedure is given as:

$$\pi_{ij} = \frac{1}{\sum_{i=1}^{N} p_i^{\alpha}\left(1-p_i^{\beta}\right)/\left(1-2\,p_i^{\beta}\right)}\left[\frac{p_i^{\alpha}\,p_j\left(1-p_i^{\beta}\right)}{\left(1-p_i\right)\left(1-2\,p_i^{\beta}\right)} + \frac{p_j^{\alpha}\,p_i\left(1-p_j^{\beta}\right)}{\left(1-p_j\right)\left(1-2\,p_j^{\beta}\right)}\right] \quad (2.2)$$

3. Some Results For New Selection Procedure

In this section we have verified some of the results for the quantities π_i and π_{ij} obtained under the new selection procedure. These results are very important for validity and applicability of a selection procedure.

Result – 1: The values of π_i and π_{ij} reduces to the standard results of simple random sampling for $p_i = p_j = \frac{1}{N}$.

Proof: To verify that π_i and π_{ij} reduces to the probability of inclusion and joint probability of inclusion of simple random sampling for $p_i = p_j = \frac{1}{N}$ we proceed as under:

$$\pi_i = \frac{N^{\alpha}\left(N^{\beta}-2\right)}{N\left(N^{\beta}-1\right)}\left[\frac{\left(N^{\beta}-1\right)\left(N-2\right)}{N^{\alpha}\left(N-1\right)\left(N^{\beta}-2\right)}+\frac{N\left(N^{\beta}-1\right)}{N^{\alpha}\left(N-1\right)\left(N^{\beta}-2\right)}\right]$$

$$\pi_i = \frac{2}{N} \tag{3.1}$$

Now

$$\pi_{ij} = \frac{1}{\sum_{i=1}^{N} p_i^{\alpha}\left(1-p_i^{\beta}\right)/\left(1-2p_i^{\beta}\right)}\left[\frac{p_i^{\alpha}\, p_j\left(1-p_i^{\beta}\right)}{\left(1-p_i\right)\left(1-2p_i^{\beta}\right)}+\frac{p_j^{\alpha}\, p_i\left(1-p_j^{\beta}\right)}{\left(1-p_j\right)\left(1-2p_j^{\beta}\right)}\right]$$

Put $P_i = 1/N$ in π_{ij}

$$\pi_{ij} = \frac{N^{\alpha}\left(N^{\beta}-2\right)}{N\left(N^{\beta}-1\right)}\left[\frac{\left(N^{\beta}-1\right)}{N^{\alpha}\left(N-1\right)\left(N^{\beta}-2\right)}+\frac{\left(N^{\beta}-1\right)}{N^{\alpha}\left(N-1\right)\left(N^{\beta}-2\right)}\right]$$

$$\pi_{ij} = \frac{2}{N\left(N-1\right)} \tag{3.2}$$

which is the joint probability of inclusion for ith and jth unit in the sample for a sample of size two in case of simple random sampling without replacement.

Result – 2: $\sum_{i=1}^{N}\pi_i = 2$ for this selection procedure.

Proof: To prove this result considers π_i as:

$$\pi_i = \frac{1}{\sum_{i=1}^{N} p_i^{\alpha}\left(1-p_i^{\beta}\right)/\left(1-2p_i^{\beta}\right)}\left[\frac{p_i^{\alpha}\left(1-p_i^{\beta}\right)\left(1-2p_i\right)}{\left(1-p_i\right)\left(1-2p_i^{\beta}\right)}+p_i\sum_{j=1}^{N}\frac{p_j^{\alpha}\left(1-p_j^{\beta}\right)}{\left(1-2p_j^{\beta}\right)\left(1-p_j\right)}\right]$$

$$\sum_{i=1}^{N}\pi_i = \frac{1}{\sum_{i=1}^{N} p_i^{\alpha}\left(1-p_i^{\beta}\right)/\left(1-2p_i^{\beta}\right)}\left[2\sum_{i=1}^{N}\frac{p_i^{\alpha}\left(1-p_i^{\beta}\right)\left(1-p_i\right)}{\left(1-p_i\right)\left(1-2p_i^{\beta}\right)}\right]$$

$$\sum_{i=1}^{N}\pi_i = 2 \tag{3.3}$$

Result – 3: The quantity π_{ij}, obtained under this selection procedure, satisfies the relation $\sum_{\substack{j=1\\ j\neq i}}^{N} \pi_{ij} = \pi_i$.

Proof: Consider π_{ij} as:

$$\pi_{ij} = \frac{1}{\sum_{i=1}^{N} p_i^{\alpha}\left(1-p_i^{\beta}\right)/\left(1-2p_i^{\beta}\right)}\left[\frac{p_i^{\alpha}\, p_j\left(1-p_i^{\beta}\right)}{\left(1-p_i\right)\left(1-2p_i^{\beta}\right)}+\frac{p_j^{\alpha}\, p_i\left(1-p_j^{\beta}\right)}{\left(1-p_j\right)\left(1-2p_j^{\beta}\right)}\right]$$

$$\sum_{j\neq i=1}^{N}\pi_{ij} = \sum_{j\neq i=1}^{N}\frac{1}{\sum_{i=1}^{N} p_i^{\alpha}\left(1-p_i^{\beta}\right)/\left(1-2p_i^{\beta}\right)}\left[\frac{p_i^{\alpha}\, p_j\left(1-p_i^{\beta}\right)}{\left(1-p_i\right)\left(1-2p_i^{\beta}\right)}+\frac{p_j^{\alpha}\, p_i\left(1-p_j^{\beta}\right)}{\left(1-p_j\right)\left(1-2p_j^{\beta}\right)}\right]$$

$$\sum_{j\neq i=1}^{N}\pi_{ij} = \frac{1}{\sum_{i=1}^{N} p_i^{\alpha}\left(1-p_i^{\beta}\right)/\left(1-2p_i^{\beta}\right)}\left[\frac{p_i^{\alpha}\left(1-p_i^{\beta}\right)}{\left(1-2p_i^{\beta}\right)}\left(\frac{1-2p_i}{1-p_i}\right)+p_i\sum_{j=1}^{N}\frac{p_j^{\alpha}\left(1-p_j^{\beta}\right)}{\left(1-p_j\right)\left(1-2p_j^{\beta}\right)}\right]$$

So

$$\sum_{j\neq i=1}^{N}\pi_{ij} = \pi_i \tag{3.4}$$

Result – 4: The quantity π_{ij}, obtained under this selection procedure, satisfies the relation $\sum_{i=1}\sum_{\substack{j=1\\ j\neq i}}^{N}\pi_{ij} = 2$ where n is the sample size.

Proof: Consider π_{ij} as:

$$\sum\sum_{j\neq i=1}^{N}\pi_{ij} = \sum_{j\neq i=1}^{N}\sum\frac{1}{\sum_{i=1}^{N} p_i^{\alpha}\left(1-p_i^{\beta}\right)/\left(1-2p_i^{\beta}\right)}\left[\frac{p_i^{\alpha}\, p_j\left(1-p_i^{\beta}\right)}{\left(1-p_i\right)\left(1-2p_i^{\beta}\right)}+\frac{p_j^{\alpha}\, p_i\left(1-p_j^{\beta}\right)}{\left(1-p_j\right)\left(1-2p_j^{\beta}\right)}\right]$$

$$\sum\sum_{j\neq i=1}^{N}\pi_{ij} = \frac{1}{\sum_{i=1}^{N} p_i^{\alpha}\left(1-p_i^{\beta}\right)/\left(1-2p_i^{\beta}\right)}\left[\sum_{i=1}^{N}\frac{p_i^{\alpha}\left(1-p_i^{\beta}\right)}{\left(1-2p_i^{\beta}\right)\left(1-p_i\right)}\left\{1-2p_i+1\right\}\right]$$

$$\sum_{j\neq i=1}^{N}\sum \pi_{ij} = \frac{2\sum_{i=1}^{N} p_i^{\alpha}\left(1-p_i^{\beta}\right)/\left(1-2\,p_i^{\beta}\right)}{\sum_{i=1}^{N} p_i^{\alpha}\left(1-p_i^{\beta}\right)/\left(1-2\,p_i^{\beta}\right)} = 2 \qquad (3.5)$$

4. Empirical Study

In this section we have given the results of the empirical study that the different pair of α and β. To carry out the empirical study we have selected 20 natural populations from the standard text on sampling techniques. The empirical study has been carried out using various values of the constant α and β in the range of -5 to 5 with an increment of 1.0. The ranking has also been done in order to decide the populations where the new selection procedure can be applied with a specified value of α and β.

Average Ranks of the Variance of the Horvitz – Thompson Estimator for the different values of α and β

α \ β	-5	-4	-3	-2	-1	1	2	3	4	5
-5	74.450	69.650	63.850	52.325	36.425	18.450	27.800	40.475	55.475	66.250
-4	74.925	69.975	64.700	52.600	37.125	18.700	27.400	39.450	54.625	65.700
-3	75.475	70.275	65.275	53.850	37.450	18.450	26.500	38.700	53.850	65.300
-2	76.000	70.875	66.050	54.350	39.700	18.250	25.350	35.300	52.300	64.750
-1	75.950	73.250	66.700	55.450	40.800	18.950	24.400	37.100	49.700	63.900
1	72.650	67.750	60.850	48.050	30.550	19.100	32.650	45.550	59.200	68.500
2	73.150	68.225	61.800	48.450	32.850	18.950	29.950	43.500	58.600	67.550
3	73.875	68.725	62.675	50.050	34.250	18.775	30.000	42.700	57.850	66.800
4	74.275	68.975	63.350	51.050	34.875	18.475	29.400	41.900	56.875	66.625
5	74.600	69.700	63.850	51.625	35.975	18.75	28.400	40.675	56.025	66.425

5. Conclusion

In the empirical study we see that the average ranks of the pair are varying from pair to pair. The pair of constant $\alpha=-2$ and $\beta=1$ gives the smallest average rank of 18.25. The pair of constant $\alpha=-3$ and $\beta=1$ gives the second smallest average rank 18.45.

From this it can be said that pair $\alpha = -2$ and $\beta = 1$ is performing best in these populations.

For the value of constant $\alpha = -2$ and $\beta = 1$ the probability of inclusion for ith unit and the joint probability of inclusion of ith and jth unit are given:

$$\pi_i = \frac{1}{\sum_{i=1}^{N} (1-p_i) / p_i^2 (1-2p_i)} \left[\frac{1}{p_i^2} + p_i \sum_{j=1}^{N} \frac{1}{p_j^2 (1-2p_j)} \right]$$

and

$$\pi_{ij} = \frac{1}{\sum_{i=1}^{N} (1-p_i) / p_i^2 (1-2p_i)} \left[\frac{p_j}{p_i^2 (1-2p_i)} + \frac{p_i}{p_j^2 (1-2p_j)} \right]$$

References:

1. Brewer, K. R. W. and Hanif, M. (1983) "*Sampling with Unequal Probabilities*", Lecture notes to Statistics, No. 15, Springer – Verlag.
2. Chudhry, A and Vos, J.W.E (1986) "Unified Theory and Strategy of Survey Sampling", North Holland Series in Statistics & Probability.
3. Hansen, M. H. and Hurwitz, W. N. (1943) "On the theory of sampling from a finite population", *Ann. Math. Stat*. 14, 333 – 362.
4. Horvitz, D. G. and Thompson, D. J. (1952) "A generalization of sampling without replacement from a finite universe", *J. Amer. Stat. Assoc*. 47, 663 – 685.
5. Madow, W. G. (1949) "On the theory of systematic sampling II", *Ann. Math. Stat.* 20, 333 – 354.
6. Yates, F. and Grundy, P. M. (1953) "Selection without replacement from within strata with probability proportional to size", *J. Roy. Stat. Soc*., B, 15, 153 – 161.